What constitutes good mathematics assessment?

Chris Sangwin
MSOR Network, University of Birmingham, Birmingham, B15 2TT.
C.J.Sangwin@bham.ac.uk

February 2, 2009

We first consider the nature and practise of mathematics before applying this to answer our question.

1 The nature of mathematics

Mathematics is a systematic method for structuring thought and arguments, which is tied closely to a coherent body of associated knowledge. Mathematics has a very long continuous history: it arose to solve practical problems of tax, accountancy and metrology in approximately 2500BCE, [6]. Such problems motivated the need for the development from arithmetic to algebra, and the problems themselves often survive in a recognizable form today. Mathematics subsequently developed into a number of sub-disciplines, including pure mathematics, applied mathematics and statistics.

1.1 Pure and applied mathematics

Applied mathematics, including statistics, solves problems relevant to the real world. In simplified form, applied mathematics proceeds by the following steps.

1. Information relevant to the problem must be identified and selected.
2. The relevant information is abstracted into a mathematical formulation. (Modelling)
3. Techniques, e.g. algebra or calculus, are applied to solve the mathematical formulation correctly.
4. The mathematical results are interpreted in terms of the original problem.

These steps will be considered further in Section 1.2 below.

Pure mathematics is concerned with the study of mathematical structures for their own sake. This includes the fundamental assumptions, e.g. the definition of "numbers", on which the discipline is built. It examines patterns, the consequences of assumptions and connections between different areas. Such patterns are often beautiful, intriguing and surprising. Pure mathematics also justifies when and how the standard algorithms can be used correctly. For example, how to solve equations of different types and how many solutions to expect. These algorithms form the core of the subject and are important cultural artifacts in their own right. Learning to using these algorithms correctly is key to progress in learning mathematics.

Rather than attempt to distinguish between pure and applied mathematics we highlight the difference between deductive and empirical justification for truth. The mathematician justifies their work deductively from stated hypotheses whereas the experimental scientist looks for empirical evidence. Therefore, mathematical *proof* of a result and *deductive justification* are central hallmarks of mathematics. The applied mathematician needs to use and acknowledge both: checking the validity of a modelling assumption against an observation of a physical system is an empirical process.

In both the pure and applied sphere, mathematicians often do not discover their results by working deductively. They use analogy, intuition, experiment and their previous experience. Hence, the final product of mathematical activity, i. e. the theorem together with it's proof, differs significantly from the process by which it is discovered. See [5] and [4].

Mathematics is unusual in the extent to which one topic builds directly upon another. Progress can be made without complete mastery of a prerequisite topic, indeed it is argued by e.g. [3] that important forms of learning take place when a technique is used as part of a more complex process. Nevertheless, one of the hallmarks of mathematics is very highly structured knowledge.

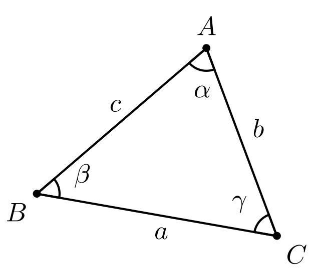
Statistics is similar to applied mathematics in that doing statistics involves: setting up a problem; planning the statistical methodology; actual data collection; presentation/analysis; followed by discussion/conclusions. In cases where the evidence is not *statistically significant* the cycle needs to be re-visited. Notice, crucially, that the statistical work precedes any data collection.

1.2 Problem solving

The first step to solving any problem is to identify the relevant information, either from the problem itself or from cultural knowledge, e.g. metrological units and physical laws. This is abstracted into a mathematical formulation, for example as algebraic equations, geometric relationships or differential equations. During the process, units and dimensional consistency provide useful checks and allow absurdities to be spotted.

When abstracting a given problem into a mathematical formulation it is important to understand and adopt appropriate conventions. For example, in algebra letters towards the beginning of the alphabet, a, b, c, are used to denote *known but as yet unspecified numbers* while those at the end x, y, z are generally used for *unknown numbers*, which represent the solution to the problem in hand. Geometry traditionally uses the following conventions:

1. *points* are upper case Roman letters, A, B, C etc.;
2. *lines* or *segments* or *curves* are lower case Roman letters, a, b, c etc.;
3. *angles* are Greek letters, α, β, γ etc.



Notice that in a triangle ABC the side opposite, a, the point A, and the angle α all correspond. Then it is immediately clear that "side a in ABC" refers to the side opposite A.

By using notation in conventional ways problems become much easier to solve: it is much easier to recognize abstract things when written in standard forms. The consistent and confident use of notation in this way is an important part of professional practise.

Once a problem has been formulated standard techniques can be applied, and this needs to be done accurately and correctly. Where there is a choice of mathematical model for a particular problem, the choices made in how the problem is formulated has a direct bearing on the difficulty of the resulting equations. Hence, in genuine modelling situations knowledge of what types of systems can be solved with the standard algorithms is key in deciding how to set up a model. Furthermore, experience with how to use these algorithms affects the precise ways in which the problems are modelled. A particular choice of coordinates can have a profound effect on the difficulty of the resulting algebra, and can make all the difference between success and failure in finding any solutions.

Historically, practical problems have provided much of the motivation for the development of the general methods of pure mathematics. For example, a physical system which is modelled by equations which cannot be solved motivates a new area of research. Once the original system has been solved, pure mathematical concerns generalize the techniques and seeks connections to apparently unrelated areas.

Once solutions have been derived, they need to be considered. Do experiments with the original system agree with solutions of the mathematical model? Do we reject some solutions as irrelevant? What do the solutions predict about the behaviour of the system?

2 Assessment of mathematics

To address our question, we shall work backward from two the important goals of mathematics education to consider good summative assessment. Summative assessment is likely to differ significantly from effective formative assessment. We do not comment here on formative or evaluative assessment.

2.1 Goals of mathematics education

Two important goals of mathematics education are

1. An understanding of the importance of pure mathematics as cultural heritage. This includes beautiful patterns, systems for logically structuring thought, surprising theorems together with the proofs, and intriguing connections.
2. Developing techniques for solving real problems, and confidence and experience in doing so.

Individuals will differ in their motivation and tastes, and so education should balance both goals. Few individuals take a serious interest pure mathematics, but many more enjoy puzzles. A much larger number of people need to solve problems. Employers often ask that employees should have

1. a good grasp of the basics;
2. a good grasp of numerical solutions;
3. a greater ability to set up a model from scratch.

2.2 Summative assessment

Overarching principle: assessment of mathematics should reflect mathematical practise.

We shall try to be more specific in highlighting important aspects of mathematical practise. These aspects will form parts of “good” assessment in mathematics.

Principle 1: mathematicians solve problems.

Mathematical practise is concerned with finding solutions to problems, whether related to practical problems (applied) or internal (pure). If assessment of students is to reflect practise then problem solving must be a key part of the assessment.

There is a spectrum of problems from entirely routine situations to those where a particular insight is applicable only in one case. Naturally, a range will be selected to test the full spectrum of understanding. In summative assessment it is not sufficient only to ask students to apply standard techniques to given equations. They are not “problems” in our sense, but practice exercises. Exercises are perfectly appropriate in the formative stages of learning, but they do not constitute the product of practise. Equally, problems should avoid special tricks.

In a good assessment problem it should be clear when the correct solution has been found through straightforward internal checks. Assessments which ask a student to verify a given solution are not problem solving and are inherently less satisfactory.

> **Principle 2:** mathematicians justify their solutions. The outcome of mathematics is a correct chain of reasoning, from agreed hypotheses to a conclusion.

Extended chains of reasoning should be assessed. Simple, routine, exercises are described by [2] as *single piece jigsaws*. It is the justification which is as important as the answer in mathematics. Indeed, in pure mathematics the answer may be "obvious", but it might be fiendishly difficult to justify.

We note that *guess and check* or *trial and improvement* are not appropriate techniques for summative assessment. In numerical analysis there are many algorithms which superficially resemble trial and improvement, by iterating towards a correct solution. The *mathematics* comes in justifying they converge and the limit to which they converge is the required solution.

> **Principle 3:** accuracy is important.

An explicit consequence of this principle is the need for students to correctly link together multiple steps of calculation and reasoning. A student who makes a mistake, however "trivial", in every other step has achieved little, if anything, of any merit. It is far better for students to achieve accurate mastery of simpler techniques then to have vague and erroneous notions from more advanced areas.

Pure mathematics does not recognize "half a proof"; there are no method marks in industry.

> **Principle 4:** standard algorithms are both useful and are important cultural artifacts in their own right.

It is important to assess (i) an understanding of when it is correct to apply such algorithms; (ii) an understanding of the details of how the algorithm works; (iii) an ability to use the algorithm accurately. Many standard algorithms can be automated, e.g. arithmetic on a calculator or more advanced operations on a computer algebra system. In order to assess (i) and (ii), it is necessary for work to be written long-hand and not for many steps to be compressed. A consequence of this is that some summative assessments should be technology free, while others may make full use of the available technology, e.g. graphical calculators.

Details of the special cases are important – this reinforces the need to assess understanding of when an algorithm is really appropriate. For example, we should confirm whether students understand why "division by zero" is forbidden as the following question illustrates.

▼ Example 1

Crititize the following argument. Suppose $a = b$ *then* $ab = a^2$*, and so* $ab - b^2 = a^2 - b^2$*. Factoring gives* $b(a-b) = (a+b)(a-b)$*. Cancelling gives* $b = a + b$*. Since* $a = b$ *we have* $b = 2b$*. Hence* $1 = 2$*.*

> **Principle 5:** conventions should be distinguished from consequences.

Using conventions allows a problem to be recognized as one to which a standard technique can be applied. This is a key step in problem solving at all levels and so it is appropriate that it forms part of any assessment. However, conventions should be distinguished from logical consequences of assumptions. If a student chooses not to follow conventions and yet is clear and correct in their reasoning then the assessment criteria should accommodate this.

3 Traditional word problems

We consider only summative assessment of mathematics. Extended project work has proved to be very difficult to assess: impersonation or plagiarism are serious practical problems which cannot be ignored. Hence, we shall assume the format is the unseen, written, timed examination. Therefore we really answer the following question.

> *What constitutes good unseen examination questions in mathematics?*

The practise of mathematics detailed above might appear hopelessly ambitious and out of touch with what is achievable in school mathematics. We need *problems* which allow *extended reasoning* where *accurate* work can be judged. The solutions should make use of *routine mathematical techniques* and where students can adopt *mathematicians' conventions*. We argue that traditional word problems possess many of the features of mathematical practise and that they can be used at many levels in schools. Hence, we argue they play an important role in "good assessment". This view is widely supported.

> I hope I shall shock a few people in asserting that the most important single task of mathematical instruction in the secondary school is to teach the setting up of equations to solve word problems. [...] And so the future engineer, when he learns in the secondary school to set up equations to solve "word problems" has a first taste of, and has an opportunity to acquire the attitude essential to, his principal professional use of mathematics. [5, Vol. I, p.g. 59]

The following is a rather contrived example (see [7, Ex 65, (44)]).

▼ Example 2

A dog starts in pursuit of a hare at a distance of 30 *of his own leaps from her. He takes* 5 *leaps while she takes* 6 *but covers as much ground in* 2 *as she in* 3*. In how many leaps of each will the hare be caught?*

Or the ubiquitous

▼ Example 3

A rectangle has length $8\mathrm{cm}$ *greater than its width. If it has an area of* $33\mathrm{cm}^2$*, find the dimensions of the rectangle.*

Interpreting such problems to derive the correct equations is far from easy: problems involving rates are particularly difficult.

▼ Example 4

Alice and Bob take 2 *hours to dig a hole together. Bob and Chris take* 3 *hours to dig the hole. Chris and Alice take* 4 *hours to dig the same hole. How long would all three of them take working together?*

The temptation is to model the work of Alice and Bob as $A+B=2$, rather than $A+B=\frac{1}{2}$. There are real difficulties in reaching a correct interpretation, and hence in moving from a word problem to a mathematical system which represents it. Consider a problem related to, but different from, 4 in which pairs of people "walk into town", rather than "dig a hole". Hence, we argue that moving from such word problems to mathematical systems constitute the beginning of mathematical modelling and hence is a valid component of pratcise. Similar conceptual difficulties occur with concentration and dilution problems. But, such problems can be practiced, and specifications in curricula can ensure tricks are not used in examinations.

Word problems immediately turn a single-step mathematical exercise into a multi-step chain of reasoning. In Example 2, let l be the number of leaps taken by the dog. The problem reduces to solving

$$l = \frac{6}{5} \times \frac{2}{3} l + 30.$$

This is a simple linear equation, but it can only be arrived at by careful work on the part of the student. In Example 3 the student is free to choose either the length or width of the rectangle as a variable. Ignoring the particular letter used for the variable, this choice results in one of two different equations, i.e. $x(x+8) = 33$ or $x(x-8) = 33$. One solution must be discarded as "unrealistic": a valuable critical judgement by the student.

Being able to select and correctly use standard techniques presupposed a certain level of fluency, which only comes with practice. Seeing the practise of mathematics itself as solving real problems through modelling, and thus understanding the satisfaction of dispatching the routine steps accurately and efficiently, may act as motivation for students to undertake the repetitive work needed for the acquisition of skills.

Problem solving is difficult. It is perfectly reasonable, and significantly challenging, to ask students to abstract information from a word problem; formulate it using conventions; recognize this as a standard case for which a known technique is applicable; and accurately solve the equations.

The use of word problems has been the subject of much research, and their use is controversial. Traditional word problems assume a certain level of cultural knowledge. Whatever the purpose of word problems, they are certainly not intended as a test of such cultural knowledge. We reasonably expect all students to be familiar with the SI system of metrology, time and currency. However, students may not be familiar with the rules of sports, or other games. Hence, when used as summative assessments due consideration needs to be given to ensure all students are treated equitably.

If word problems are "abstract" and decontextualized then they appear contrived, divorced from reality and even ridiculous. However, the analysis of [1] found that social class was a significant factor in determining childrens performance. In particular they report that *"working class students performed equally as well as their middle class counterparts on decontextualised test items but struggled on realistic items which were embedded in everyday contexts"*.

Word problem have many features which correspond to the practise of mathematics. Their use in summative mathematics examinations requires care to ensure that mathematical practise, rather than cultural or social background, is really the focus of assessment. Hence, care is needed, as ever, with the precise details of how the problems themselves are formulated and used.

References

[1] B. Cooper and Dunne M. *Assessing childrens mathematical knowledge*. Open University Press, 2000.

[2] T. Gardiner. Beyond the soup kitchen: Thoughts on revising the mathematics "strategies/frameworks" for england. *International Journal for Mathematics Teaching and Learning*, May 2006. ISSN 1473 - 0111.

[3] D. Hewitt. Mathematical fluency: the nature of practice and the role of subordination. *For the learning of mathematics*, 16(2):28–35, 1996.

[4] I. Lakatos. *Proofs and refutations*. Cambridge University Press, 1976.

[5] G. Polya. *Mathematical discovery: on understanding, learning, and teaching problem solving*. Wiley, 1962.

[6] E. Robson. *Mathematics in Ancient Iraq*. Princeton University Press, 2008.

[7] C. O. Tuckey. *Examples in Algebra*. Bell & Sons, London, 1904.